

A STANDARDIZED PSYCHIATRIC INTERVIEW FOR USE IN COMMUNITY SURVEYS

D. P. GOLDBERG, M.R.C.P., D.P.M.

Senior Lecturer in Psychiatry, University of Manchester

B. COOPER, M.D., D.P.M.

Senior Lecturer, Institute of Psychiatry, London

M. R. EASTWOOD, M.D., D.P.M.

Senior Lecturer in Psychiatry, University of Tasmania

H. B. KEDWARD, M.D., D.P.M.

Associate Professor of Medicine, Memorial University of Newfoundland, Canada

M. SHEPHERD, D.M., M.R.C.P., D.P.M.

Professor of Epidemiological Psychiatry, Institute of Psychiatry, London

SURVEYS of psychiatric illness in the community are at present handicapped by the lack of valid and reliable methods of case-identification. The widely varying estimates of psychiatric prevalence made by different workers in this field (Lin and Standley, 1962) emphasize the urgent need for such techniques, which could be used both in field surveys and in the screening of general practice populations.

In large-scale psychiatric surveys, the use of a two-stage screening procedure is desirable and may, indeed, be essential for economic reasons. The first stage entails the selection of possible or 'potential' cases by means of a rapid and simply administered screening test; the second comprises detailed clinical examination of such potential cases in order that they may be confirmed as actual cases (Blum, 1962) and given an accurate diagnostic assessment. The present paper is concerned only with the second of these stages, namely, the development of a standardized psychiatric interview and rating technique suitable for application to potential cases in a community setting.

A number of standardized psychiatric interviews, including some which are highly reliable, have been developed in recent years, but for various reasons all are unsuitable for field surveys. In the United States, Lorr, Klett, and McNair (1963) have described an assessment based on the Inpatient Multidimensional Psychiatric Scale (IMPS), and Overall and Gorham (1962) have used a psychiatric assessment derived from a shortened version of the Lorr scale; both these scales heavily emphasize psychotic

phenomena which are relatively uncommon in the general population. The interview described by Spitzer, Fleiss, Burdock, and Hardesty (1964), although in some ways more suitable, is still insufficiently flexible and contains many items which would make it unacceptable to normal individuals. In this country, Wing, Birley, Cooper, Graham, and Isaacs (1967) and Kendell, Everitt, Cooper, Sartorius, and David (1968) have published accounts of the 'Present State Examination' which has been designed primarily for use in international studies. This very comprehensive 500-item schedule was designed for administration to known psychiatric patients and again does not readily lend itself in its present form for use in community surveys.

With these considerations in mind, a standardized psychiatric interview has been constructed to meet the following requirements:

- (1) Psychiatric assessment should be made by an experienced psychiatrist in a realistic clinical setting;
- (2) The interview should be acceptable to individuals who may not see themselves as psychiatrically disturbed;
- (3) The content of the interview should be appropriate to the types of psychiatric disturbance commonly encountered in the community;
- (4) The interview should generate information about individual symptoms and signs of illness as well as an overall diagnostic assessment;

Reprint requests to M.S., Institute of Psychiatry, de Crespigny Park, Denmark Hill, London, S.E.5

(5) It should discriminate clearly both between mentally disturbed and normal individuals and between patients with different degrees of psychiatric disturbance;

(6) It should be relatively economical of time so that large numbers of patients can be included:

(7) The psychiatric assessments and clinical ratings should be reliable in the sense of being reproducible by different trained observers.

A final point concerns the validity of assessments based on the interview procedure. Since for most psychiatric disorders there are no objective standards against which clinical measurements can be compared, examination by an experienced psychiatrist remains the best available method of case-identification. Such expert judgments will be more authoritative if substantial agreement can be shown to exist between two or more clinicians working independently, but then the issue becomes one of reliability rather than of validity. Hence, while the 'content validity' (Cronbach, 1949) of the interview in question is undoubtedly satisfactory, the best measure of its quality as a research tool must be its reliability when administered by trained psychiatrists. The establishment of this inter-rate reliability was accordingly given first priority in the present study.

DESIGN OF THE INTERVIEW SCHEDULE*

The interview schedule is divided into four sections. The first is unstructured and consists essentially of sub-headings for brief recording of the patient's present and past medical history. This part of the interview, which normally takes only a few minutes to complete, is used to establish rapport and to put the patient at ease.

The second part of the interview is a more detailed and systematic enquiry about any psychiatric symptoms which the patient may have experienced in the preceding week. The symptoms are arranged in 10 groups which are administered to the patient in the following order:

- Somatic symptoms
- Fatigue
- Sleep disturbance
- Irritability
- Lack of concentration
- Depression
- Anxiety and worry
- Phobias
- Obsessions and compulsions
- Depersonalization

In this context, the term 'somatic symptoms' is meant to include any bodily symptom which, in

the interviewer's opinion, has been precipitated by psychological events, or which is being maintained or prolonged by psychological factors. Inevitably there are many cases where, even after detailed questioning, the interviewer is still in doubt about whether a particular symptom should be rated under this heading, and in these cases the symptom is not included.

The ordering of the symptoms was designed to provide a progression from those which are commonly encountered in everyday life (minor aches and pains, fatigue, loss of sleep) to those which are more obviously morbid mental phenomena. The interviewer does not ask about delusions and hallucinations unless the patient's responses to other questions lead him to suspect a psychotic illness. In the same way, cognitive and memory testing is applied only when there is reason to suspect intellectual impairment. The interviewer is free to pose any additional questions which may be indicated in a given case, his aim being to establish the frequency, duration, and intensity of each symptom during the preceding week. There are simple rules relating the frequency and intensity of each symptom to the 5-point scale on which it is rated. Figure 1 illustrates a typical specimen page from this section of the schedule, that dealing with 'Irritability'. The mandatory question appears above the horizontal line.

Do you find that you are easily upset or irritable with those around you?

If the patient's reply indicates irritability, go on as follows:

How long have you been like this?
Are you like it all the time, or just occasionally?
What sort of things upset you?
How has it been in the past week?
Have you had any rows with anyone in the past week?
Are there still any hard feelings?
IRRITABILITY 4 3 2 1 0

FIG. 1. A sample page from the Clinical Schedule.

The third section of the interview schedule is unstructured and permits the interviewer to collect just as much information about the patient's family and personal history as he feels necessary to make his clinical assessment.

As soon as the interview is over and the patient has left the room, the interviewer completes the fourth and final section of the schedule in which he rates the abnormalities observed during the interview

*Further information about the schedule may be obtained from the General Practice Research Unit, Institute of Psychiatry, London, S.E.5

on 12 5-point scales. These ratings represent the doctor's view of the 'manifest abnormalities' that he has observed at interview, as distinct from the patient's symptoms that have already been rated in the second part of the interview. There are three scales for abnormalities of behaviour, four for abnormal moods, and five for perceptual and cognitive abnormalities:

Slow, lacking spontaneity
 Suspicious, defensive
 Histrionic
 Depressed
 Anxious, tense
 Elated, euphoric
 Flattened, incongruous
 Excessive concern with bodily functions
 Depressive thought content
 Thought disorder, delusions, misinterpretations
 Hallucinations
 Intellectual impairment

From the standpoint of clinical psychiatry, this list is unusual in terms of balance and omissions. It was, however, arrived at only after lengthy preliminary work as the best operational framework to be adopted in order to achieve the aims set out in the introduction. Some items were abandoned because raters were unable to achieve satisfactory agreement about them, while items dealing with psychotic phenomena were condensed into three or four groups because of the comparative rarity of psychotic illness in a community setting. The

grouping of these items was based on the underlying assumption that distinct or overlapping clinical phenomena can be rated on a single scale; for example, psychotic agitation on the same scale as neurotic tension, and thought disorder on the same scale as misinterpretations or over-valued ideas.

The interviewer makes in all a total of 22 ratings, 10 for reported symptoms in the past week and 12 for abnormalities manifest at interview. Each rating is made on a 5-point scale as shown in Fig. 1. In broad terms a rating of 0 indicates absence of a symptom; 1 indicates an habitual trait or borderline symptom which does not cause significant distress or require treatment; 2, 3, and 4 indicate respectively mild, moderate, and severe degrees of clinical severity of a definitely morbid symptom.

At an early stage of the work it became apparent that to achieve satisfactory reliability all operational terms would have to be carefully defined; these working definitions, together with detailed criteria for rating, were eventually gathered together in a manual which can be consulted on any doubtful points that may arise.

This clinical manual contains general guidance on the conduct of the interview and the way of using the 5-point scales for the various ratings; in addition it contains a page for each reported symptom and manifest abnormality giving detailed guidance on the definition of each item and the use of the 5-point scale for that item. Figure 2 shows the page in the manual which corresponds to the page in the

IRRITABILITY

The term is used to cover both irritability and hypersensitivity in personal relationships. It should only be extended to cover inward brooding if this has affected the respondent's relationships with others. The term should apply only to behaviour which has been noticeable to others in the respondent's environment. In assessing any distress that the respondent may have caused others the rater must rely on what the patient says, and must *not* extrapolate from other features of the history or the mental state in order to make his assessment.

Rating '0' Absent: Negative answer to the mandatory question.

Rating '1' Mild: While irritability may have been present in a minor degree, it either represents an habitual trait which has not caused significant distress, OR from what the respondent says the rater thinks that irritability has not caused significant distress to either the respondent or to others.

Rating '2' Moderate: Irritability has occurred occasionally in the past week in a degree sufficient to cause significant distress to either the patient or to those in his environment.

Rating '3' Marked: Irritability as defined in '2' above has been present frequently throughout the past week

OR

In a context of increased irritability there has been a single major row or flare-up with those in his environment.

Rating '4' Severe: In a context of increased irritability there have been a series of major rows or flare-ups. This rating should also be used for single rows where physical violence has been committed by the respondent.

FIG. 2. A sample page from the Clinical Manual dealing with irritability.

Interview Schedule dealing with irritability.

Finally, the interviewer writes a short diagnostic formulation and, if appropriate, makes a psychiatric diagnosis drawn from the I.C.D. (General Register Office, 1968). The whole procedure can take from as little as 10 minutes for a healthy individual with no medical complaints to over an hour for a poly-symptomatic patient.

THE RELIABILITY STUDY

Although the interview was designed for use in a community setting, the initial reliability study had to be undertaken on patients with many symptoms of mental illness, so as to test the instrument on the widest possible range of psychiatric phenomena. For this reason, the patients selected for the study were all hospital in-patients or day-patients.*

Once the structure of the interview had been decided, a small pilot study was carried out with four raters and 12 patients. Each patient was seen by two raters, one of whom administered the interview while the other acted as co-rater. The patients were randomly assigned to pairs of raters and within the pairs each psychiatrist acted as interviewer and once as co-rater. From the experience gained in this pilot study, a number of alterations were made to the interview schedule and the accompanying manual.

The main reliability study was carried out on a total of 40 patients, divided into two sets of 20 each. In each set five psychiatrists participated, the experimental design being similar to that described for the pilot study. With one replacement for the second set of interviews, the raters were the same throughout.

Most of the standardized psychiatric interviews mentioned in the introductory section have been assessed for reliability in terms of product-moment correlations between observers, computed for each item scale. The analysis carried out by Kendell *et al.* (1968) was more thorough in that they used a three-way analysis of variance (the dimensions being doctors, patients, and symptoms) and also computed for each of their items a statistic called 'weighted kappa', originally proposed by Cohen (1960, 1969) and recently discussed in some detail by Everitt (1968). 'Weighted kappa' is a reliability coefficient between observers that takes account of the relative seriousness of the different types of disagreement that can occur; it has a value of 1.0 if agreement is perfect, of zero if agreement is no better

than that due to chance, and a negative value if agreement is less than chance.

To allow direct comparison with other published work, all these analyses have been carried out in the present study. Table I shows the product-moment correlation and weighted kappa value for each item separately. It can be seen that the correlations are of satisfactory magnitude for all items with the exception of 'histrionic'.

TABLE I
REPORTED SYMPTOMS AND MANIFEST ABNORMALITIES
RATED AT INTERVIEW, SHOWING RELIABILITY CO-
EFFICIENTS BETWEEN INTERVIEWER AND CO-RATER
FOR EACH ITEM

Item	Product-Moment Correlation	Weighted Kappa
Reported Symptoms:		
Somatic symptoms	0.787	0.6733
Fatigue	0.805	0.7654
Sleep disturbance	0.981	0.8013
Irritability	0.841	0.6761
Lack of concentration	0.861	0.6745
Depression	0.914	0.8000
Anxiety and worry	0.830	0.6699
Phobias	0.789	0.7391
Obsessions and compulsions	0.837	0.6696
Depersonalization	0.870	0.7143
Manifest Abnormalities		
Slow, lacking spontaneity	0.913	0.6000
Suspicious, defensive	0.858*	0.7368*
Histrionic	0.664*	0.4828*
Depressed	0.902	0.6646
Anxious, tense	0.773	0.6117
Elated, euphoric	0.981*	0.9362
Flattened, incongruous	0.804	0.724
Depressed in thought content	0.766	0.6501
Excessive concern with bodily functions	0.829	0.7510
Thought disorder, delusions, misinterpretations	0.832	0.7113
Hallucinations	0.956	0.8789
Intellectual impairment	0.874	0.7753

*Based on 20 patients only

The analysis of variance, a 3-way model without replications, is shown in Table II. For the purposes of this analysis, reported symptoms and manifest abnormalities have been grouped together simply as 'items'.

TABLE II
RESULTS OF THE 3-WAY ANALYSIS OF VARIANCE

	SSq.	d.f.	MSq.	
Source of Variations				
Between patients	83.54	39	2.14	**
Between items	74.47	19	3.92	**
Between doctors	0.12	1	0.12	NS
Interactions				
Patients × items	259.37	741	0.35	*
Patients × doctors	3.50	39	0.09	NS
Items × doctors	2.10	19	0.11	NS
Residual				
Patients × doctors × items	148.52	741	0.20	
Total	571.61	1599		

*At the Bethlem Royal and Maudsley Hospitals, or St. Francis Hospital, East Dulwich

**Significant at 0.01 level

*Significant at 0.05 level

When the mean squares for the main variables—doctors, patients, and items—were tested against the mean square for the residual it was found that the mean squares for patients and items were significant at the 0.01 level, but that the mean square for doctors was not. This indicates that while there was significant variation between patients and between items there was not significant variation between interviewer and co-rater. The significant interaction between patients and items merely indicates the tendency for different patients to have different symptoms. On the other hand, the small mean square values for interaction between doctors and items and between doctors and patients indicate that the doctors' rating standards remained consistent from one item to the next and from one patient to the next.

Maxwell and Pilliner (1968) have described a technique for using a three-way analysis of variance to estimate reliability between raters by comparing the between-patient mean square with that for interaction between patients and doctors. Using this method on the present data, the reliability coefficient between raters was calculated as +0.919.

DISCUSSION

The overall standard of reliability for this interview compares favourably with the results published by other workers; for example, the 'weighted kappa' values for reported symptoms are rather higher than those obtained by Kendell *et al.* (1968). The reliability for the 22 separate item scales was satisfactory with the single exception of 'histrionic' as a description of interview behaviour. The poor reliability of this item probably relates to the fact that it is at least as much a personality variable as an index of morbidity. It is, however, of some importance in that clinical psychiatrists habitually take account of what they regard as affected or role-playing behaviour in evaluating a patient's reported symptoms.

The interview was found to meet its requirements in being relatively economical of time, easy to administer, and flexible in that it can be adapted to the purposes of various types of enquiry. Perhaps most important, it has proved acceptable to a wide range of patients. So far, it has been administered to over 400 general practice surgery attenders and nearly as many individuals seen in a community screening programme; refusals and adverse responses have been very few. Moreover, the clinicians who have administered the interview have found it appropriate for assessing and rating a wide range of psychiatric morbidity.

A number of further studies will be required for complete evaluation of the interview, in particular, its reliability when administered by other psychiatrists and to different groups of subjects. The high reliability found in the present study derives partly from the common background and training of the participants, and partly from preliminary training and careful use of the manual of definitions. Being only partially structured, the interview allows the psychiatrist who uses it a good deal of freedom; consequently, it would not be reliable in inexperienced hands.

The individual ratings can be treated in several ways. Scores for each patient constitute a symptom profile which can serve as an aid to diagnosis. This has not been attempted in the present reliability study, but it will be the subject of a future study which will seek to relate the symptoms profile to I.C.D. diagnoses. The 22 ratings for each patient can be summed and weighted to give an overall index of clinical severity. In the present stage of knowledge any such weighting procedure must be arbitrary; but for the present series, the best agreement with clinical judgment was obtained by taking the sum of the symptom ratings plus twice the sum of the manifest abnormality ratings.

The overall severity score, S , can therefore be obtained from the simple formula:

$$S = \Sigma R_i + 2 \Sigma M_i$$

where R_1, R_2, \dots, R_{10} = Reported symptom ratings

M_1, M_2, \dots, M_{12} = Manifest abnormality ratings.

An alternative way of using the scores is to compare the symptom profiles of two or more groups of patients. The mean score for each item is computed for each group; the data can then be assessed by using a 't-test' to compare the means for the two groups for each separate item, or by constructing a histogram showing the mean symptom profiles for each group.

As a research tool, the interview has a number of possible uses. First, it could be employed in large-scale community surveys as the second stage of the case-finding procedure. From this point of view it represents an advance over the second-hand methods of case-identification which have been widely used in psychiatric field surveys (see, *inter alia*, Srole, Langner, Michael, Opler, and Rennie, 1962; Leighton, Harding, Macklin, Macmillan, and Leighton, 1963).

Secondly, the interview could be administered to all members of a defined population sample in seek-

ing to test for associations between psychiatric disturbance and clinical or social variables, for example, in the comparison of demographically matched groups known to differ in respect of physical morbidity.

Thirdly, the interview could be used to establish change in the psychiatric state of patients over a known time interval, as in charting the course of different clinical syndromes, or in the evaluation of a therapeutic regime. In one study carried out by the General Practice Research Unit, a group of 97 patients with psychiatric illnesses have been interviewed with the clinical schedule on two occasions six months apart, and large changes were demonstrated in the symptom profiles of those patients whose clinical status was thought to have altered.

Finally, it could be employed in the comparison of population samples drawn from different areas for examining differences in both symptomatology and prevalence. Cross-national studies would present special problems but one preliminary study of patients in London and Newfoundland (Goldberg and Kedward, 1970) suggests that the difficulties could be overcome.

SUMMARY

A standardized psychiatric interview has been described which is suitable for use in community surveys and which is geared to the measurement of change in the mental state. It must be administered by an experienced psychiatrist and requires special training.

The interview can be administered to a wide range of persons and is fairly economical of time, the range being from 10 to 20 minutes for normal healthy individuals to 30 to 60 minutes for patients with frank psychiatric illness.

The reliability of the interview was measured in an experiment in which five psychiatrists interviewed two separate groups of 20 hospital patients, a co-rater being present at each interview. The individual items in the interview were found to have high reliability coefficients, comparable to those reported for other standardized interviews designed for use with identified psychiatric cases. The overall reliability coefficient between interviewer and co-rater for the whole interview was found to be +0.92.

The possible applications of this interview as a research instrument are discussed.

We are particularly grateful to Dr. D. H. Gath (now at The Warneford Hospital, Oxford) for participating in the final set of 20 interviews after one of us (H.B.K.) had left the Unit. We are indebted to those consultants at the Bethlem and Maudsley Hospitals who allowed us

to interview their patients. Our thanks are due to Professor A. E. Maxwell and his staff at the Department of Biometrics of the Institute, who gave statistical advice at all stages of the research and carried out the three-way analysis of variance and 'weighted kappa' analysis using the University of London's 'Atlas' Computer. We are also grateful to Mr. F. Gattoni of the London School of Economics for his advice and help with the analysis of data. This work was undertaken as part of a research programme supported by a grant from the Department of Health and Social Security, and by the support of one author (B.C.) by the Mental Health Research Fund.

REFERENCES

- BLUM, R. H. (1962). Case identification in psychiatric epidemiology: methods and problems. *Milbank mem. Fd. Quart.*, **40**, 253.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, **20**, 37.
- (1969). Weighted kappa: nominal scale agreement with provision for degrees of disagreement. *Amer. Psychol.* (in press).
- CRONBACH, L. J. (1949). *Essentials of Psychological Testing*. Harper, New York.
- EVERITT, B. S. (1968). Moments of the statistics kappa and weighted kappa. *Brit. J. math. stat. Psychol.*, **21**, 97.
- GENERAL REGISTER OFFICE (1968). A Glossary of Mental Disorders. Studies on Medical and Population Subjects No. 22. London, H.M.S.O.
- GOLDBERG, D. P., and KEDWARD, H. B. (1970). The social correlates of chronic psychiatric illness in Newfoundland. I. Clinical findings. In preparation.
- KENDELL, R. E., EVERITT, B., COOPER, J. E., SARTORIUS, N., and DAVID, M. E. (1968). The reliability of the 'Present State Examination'. *Soc. Psychiat.*, **3**, 123.
- LEIGHTON, D. C., HARDING, J. S., MACKLIN, D. B., MACMILLAN, A. M., and LEIGHTON, A. H. (1963). *The Character of Danger*. Basic Books, New York.
- LIN, TSUNG-YI and STANDLEY, C. C. (1962). The Scope of Epidemiology in Psychiatry. W.H.O. Public Health Papers No. 16.
- LORR, M., KLETT, C. J., and MCNAIR, D. M. (1963). *Syndromes of Psychosis*. Pergamon, London.
- MAXWELL, A. E., and PILLNER, A. E. G. (1968). Deriving coefficients of reliability and agreement for ratings. *Brit. J. math. stat. Psychol.*, **21**, 105.
- OVERALL, J. E., and GORHAM, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychol. Reports*, **10**, 799.
- SPITZER, R. L., FLEISS, J. L., BURDOCK, E. I., and HARDESTY, A. S. (1964). The Mental Status Schedule: rationale, reliability and validity. *Comprehens. Psychiat.*, **5**, 384.
- SROLE, L., LANGNER, T. S., MICHAEL, S. T., OPLER, M. K., and RENNIE, T. A. C. (1962). *Mental Health in the Metropolis: the Midtown Manhattan Study*. McGraw-Hill, New York.
- WING, J. K., BIRLEY, J. L. T., COOPER, J. E., GRAHAM, P., and ISAACS, A. D. (1967). Reliability of a procedure for measuring and classifying 'present psychiatric state'. *Brit. J. Psychiat.*, **113**, 499.